(2)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1301 May, 1991

Limitations of Non Model-Based Recognition Schemes

Yael Moses[1] and Shimon Ullman

Abstract: Different approaches to visual object recognition can be divided into two general classes: model-based vs. non model-based schemes. In this paper we establish some limitation on the class of non model-based recognition schemes. A non model-based scheme is based on functions invariant to viewing position and illumination conditions. We show that every function that is invariant to viewing position of all objects is the trivial (constant) function. The same result holds even if the recognition function is not required to be perfect, but is allowed to make mistakes and misidentify each object from a substantial fraction of viewing directions. It follows that every consistent recognition scheme for recognising 3-D objects must in general be model based.

We then consider recognition schemes restricted to classes of objects and show that, for some classes, the only consistent recognition function is still the trivial function. For other classes (such as the class of symmetric objects) a nontrivial recognition scheme exists. We define the notion of a discrimination power of a consistent recognition function for a class of objects. The function's discrimination power determines the set of objects that can be discriminated by the recognition function. We show that it is possible to determine the upper bound of the function's discrimination power for every consistent recognition function.

Acknowledgments: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by an Office of Naval Research University Research Initiative grant under contract N00014-86-K-0685, and in part by the Advanced Research Projects Agency of the Department of Defense under Army contract number DACA76-85-C-0010 and under Office of Naval Research contract N00014-85-k-0124. S.U was also supported by NFS grant IRI-8900267.

[1]Department of Applied Math., the Weizmann Institute of Science, Rehovot, Israel

Limitations of Non Model-Based Recognition Schemes

1 Introduction

Recognizing objects in images is one of the most important aspects of visual perception. From the computational point of view, object recognition is also one of the most difficult problems. The difficulties are due to the fact that different images of the same object may be very dissimilar. The image of an object depends not only on the object's shape, but also on the viewing position, the illumination conditions, other objects in the scene, and noise.

Several approaches have been proposed to deal with this problem of the dissimilarity between different views of the same object. In general, it is possible to classify these approaches into *model-based* vs. *non model-based* schemes. In this paper we examine the limitations of non model-based recognition schemes.

Let us begin with some basic definitions. A *recognition function* is a function from 2-D images to a space with an equivalence relation. Without loss of generality we can assume that the range of the function is the real numbers, R. We define a *consistent recognition function* for a set of objects to be a recognition function that has identical value on all images of the same object from the set. That is, let s be the set of objects that f has to recognize. If v_1 and v_2 are two images of the same object from the set s then $f(v_1) = f(v_2)$.

A *recognition scheme* is a general scheme for constructing recognition functions for particular sets of objects. It can be regarded as a function from sets of 3-D objects, to the space of recognition functions. That is, given a set of objects, s, the recognition scheme, g, produces a recognition function, $g(s) = f$. The *scope* of the recognition scheme is the set of all the objects that the scheme may be required to recognize. In general, it may be the set of all possible 3-D objects. In other cases, the scope may be limited, e.g., to 2-D objects, or to faces, or to the set of symmetric objects. A set s of objects is then selected from the scope and presented to the recognition scheme. The scheme g then returns a recognition function f for the set s. A recognition scheme is considered consistent if $g(s) = f$ is consistent on s as defined above, for every set s from the scheme's scope.

A model-based scheme produces a recognition function $g(s) = f$ that depends on the set of models. That is, there exist two sets s_1 and s_2 such that $g(s_1) \neq g(s_2)$ where the inequality is a function inequality. Note that the definition of model-based scheme in our discussion is quite broad, it does not specify the type of models or how they are used.

The nonlinear *RBF* interpolation scheme (Poggio & Edelman 1990) is an example of a model-based recognition scheme. In this scheme an image is represented as a point in some high dimensional space. A model is constructed from a set of images of a given object. The model defines a nonlinear subspace which is an interpolation of the model images in the high dimensional space. Given a new image, the recognition scheme computes whether the image belongs to one of the known subspaces. In this scheme, the subspaces that the function computes depend on the set of objects that have to be recognized and the images that have been presented, therefore the function value for a given image depends on the set of objects learned by the scheme. The schemes developed by Brooks (1981), Bolles & Cain (1982), Grimson & Lozano-Pérez (1984),Grimson & Lozano-Pérez (1987), Lowe (1985), Huttenlocher & Ullman (1987) and Ullman (1989) are also examples of model-based recognition schemes.

A non model-based recognition scheme produces a recognition function $g(s) = f$ that does not depend on the set of models. That is, if g is a non model-based recognition scheme, then for every two sets s_1 and s_2, $g(s_1) = g(s_2)$, where the equality is a function equality.

Non model-based approaches have been used, for example, for face recognition. In this case the scope of the recognition scheme is limited to faces. These schemes use certain relations between facial features to uniquely determine the identity of a face (Kanade 1977, Cannon *et al.* 1986, Wong *et al.* 1989). In these schemes, the relations between the facial features used for the recognition do not change when a new face is learned by the system. Other examples are schemes for recognizing planar curves (Lin 1987, Cyganski *et al.* 1987).

In this paper we consider the limitations of non model-based recognition schemes. A consistent non model-based recognition scheme produces the same function for every set of models. Therefore, the recognition function must be consistent on every possible set of objects within the schem's scope. Such a function is *universally consistent*, that is, consistent for objects in its scope.

A consistent recognition function of the set s should be invariant to at least two types of manipulations: changes in viewing position, and changes in the illumination conditions. We first examine the limitation of non model-based schemes with respect to viewing position, and then to illumination conditions.

In examining the effects of viewing position, we will consider objects consisting of a discrete set of 3-D points. The domain of the recognition function consists of all binary images resulting from scaling of orthographic projection of such discrete objects on the plane. We show (Section 2) that every consistent universal recognition function with respect to viewing position must be trivial, i.e. a constant function[1]. Such a function

[1]A similar result has been independently proved by Burns *et al.* 1990 and Clemens & Jacobs 1990.

does not make any distinctions between objects, and therefore cannot be used for object recognition. On the other hand, we show (Section 6) that in a model-based scheme it is usually possible to define a nontrivial consistent recognition function.

The human visual system, in some cases, misidentifies an object from certain viewing positions. We therefore consider recognition functions that are not perfectly consistent. Such a recognition function can be inconsistent for some images of objects taken from specific viewing positions. In Section 3.1 we show that such a function must still be constant, even if it is inconsistent for a large number of images (we define later what we consider "large"). We also consider (Section 3.2) imperfect recognition functions where the values of the function on images of a given object may vary, but must lie within a certain interval.

Many recognition schemes deal with a limited scope of objects such as cars, faces or industrial parts. In this case, the scheme must recognize only objects from a specific class (possibly infinite) of objects. For such schemes, the question arises of whether there exists a non-trivial consistent function for objects from the scheme's scope. The function can have in this case arbitrary values for images of objects that do not belong to the class. The existence of a nontrivial consistent function for a specific class of objects depends on the particular class in question. In Section (4) we discuss the existence of consistent recognition function with respect to viewing position for specific classes of objects. In Section (4.1) we give an example of a class of objects for which every consistent function is still a constant function. In Section (4.2) we discuss infinite classes of objects for which a nontrivial consistent function does exist. In Section (4.2) we also define the notion of the function *discrimination power*. The function discrimination power determines the set of objects that can be discriminated by a recognition scheme. We show that, given a class of objects, it is possible to determine an upper bound for the discrimination power of any consistent function for that class. We use as an example the class of symmetric objects (Section 4.2.2).

Finally, we consider grey level images of objects that consist of n small surface patches in space (this can be thought of as sampling an object at n different points). We show that every consistent function with respect to illumination conditions and viewing position defined on points of the grey level image is also a constant function.

We conclude that every consistent recognition scheme for 3-D objects must depend strongly on the set of objects learned by the system. That is, a general consistent recognition scheme (a scheme that is not limited to a specific class of objects) must be model-based. In particular, the invariant approach cannot be applied to arbitrary 3-D objects viewed from arbitrary viewing positions. However, a consistent recognition function can be defined for non model-based schemes restricted to specific class of objects. An upper bound for the discrimination power of any consistent recognition function can be determined for every class of objects.

It is worth noting here the differences between the existence of features that are invariant to viewing position or illumination condition, and the existence of consistent recognition functions. An example of invariant features are parallel lines that are invariant to viewing position (under orthographic projection). Other examples of features invariant to viewing position can be found in Verri & Yuille (1986) and Ponce *et al.* (1987). However, a function that merely detects invariant features (without recognizing different objects) can be regarded as a consistent recognition function that must recognize only a given set of objects, and the general results are not applicable to this case (see Section 6).

Many applications of functions that are invariant to viewing position for images of 2-D objects can be found in the literature, such as, Fourier transform invariances (Lin 1987) or moment invariances (Hu 1961, Hu 1962, and Khotanzad & Hong 1990). Our analysis applies to the general case of 3-D objects and hence does not contradict these results. The analysis of the class of 2-D objects (Lamdan & Wolfson) shows that for 2-D (pointwise) objects a non-trivial invariant function exists. Therefore, a non model-based recognition scheme with limited scope of 2-D point objects can be defined.

2 Consistent recognition function with respect to viewing position

We begin with the general case of a universally consistent recognition function with respect to viewing position, i.e. a function invariant to viewing position of all possible objects. The function is assumed to be defined on the orthographic projection of objects that consist of points in space.

Claim 1: Every function that is invariant to viewing position of all possible objects is a constant function.

Proof: A function that is invariant to viewing position by definition yields the same value for all images of a given object. Clearly, if two objects have a common orthographic projection, then the function must have the same value for all images of these two objects.

We define a *reachable sequence* to be a sequence of objects such that each two successive objects in the sequence have a common orthographic projection. The function must have the same value for all images of objects in a reachable sequence. A *reachable object* from a given object is defined to be an object such that there exists a reachable sequence starting at the given object and ending at the reachable object. Clearly, the value of the function is identical for all images of objects that are reachable from a single object.

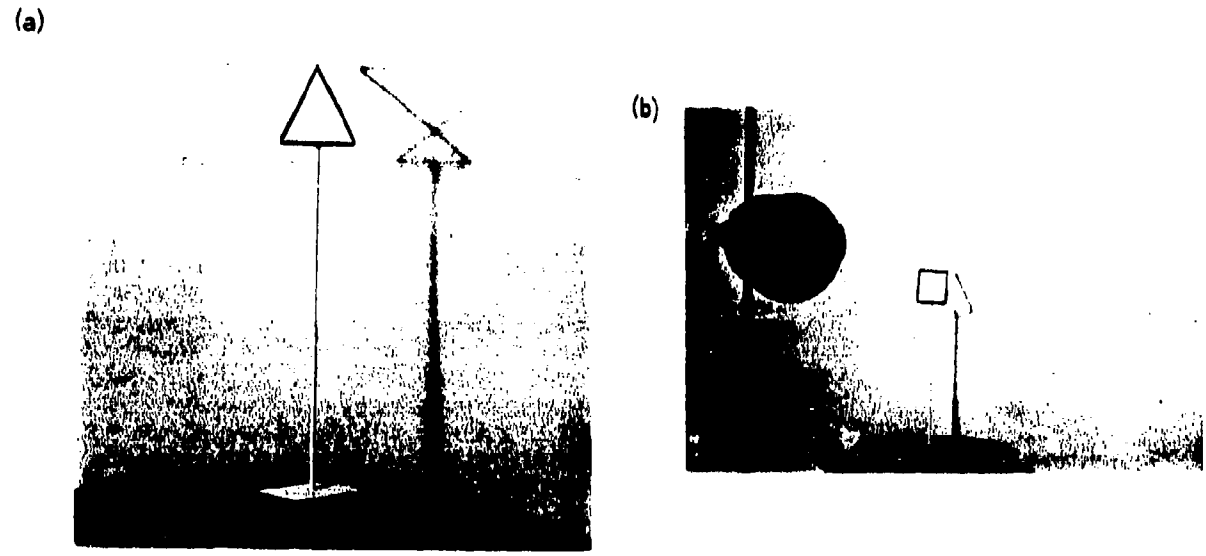

Figure 1: Two perspective views of the same three-dimensional object. (a) The object viewed from 0 deg. (b) the object viewed from 5 deg about its vertical axis.

Every image is an orthographic projection of some 3-D object. In order to prove that the function is constant on all possible images, all that is left to show is that every two objects are reachable from one another. This is shown in Appendix 1. □

An example that demonstrates how a box-like object is reachable from a pyramid was given by Ullman (1977) (see Fig 1). A wire object is shown whose projection in one direction is a triangle, and in another direction (see the shadows in Fig 1), only 5 degrees apart, is a square. Hence, any invariant function must have the same value for a box and a pyramid.

We have shown that any universal and consistent recognition function is a constant function. Any non model-based recognition scheme with a universal scope is subject to the same limitation, since such a scheme is required to be consistent on all the objects in its scope. Hence, any non model-based recognition scheme with a universal scope cannot discriminate between any two objects.

3 Imperfect recognition functions

Up to now, we have assumed that the recognition function must be entirely consistent. That is, it must have exactly the same value for all possible images of the same objects. However, a recognition scheme may be allowed to make errors. We turn next to examine recognition functions that are less than perfect. In Section ?? we consider consistent functions with respect to viewing position that can have errors on a significant subset of images. In Section ?? we discuss functions that are almost consistent with respect to viewing position, in the sense that the function values for images of the same object are not necessarily identical, but only lie within a certain range of values.

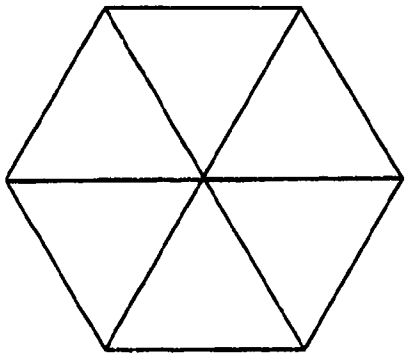

Figure 2: A cube perceived as a 2-D hexagonal when viewed from a certain viewing position.

3.1 Errors of the recognition function

The human visual system may fail in some cases to identify correctly a given object when viewed from certain viewing positions. For example, it might identify a cube from a certain viewing angle as a 2-D hexagon (Fig 2). The recognition function used by the human visual system is inconsistent for some images of the cube. The question is whether there exists a nontrivial universally consistent function, when the requirements are relaxed: for each object the recognition function is allowed to make errors (some arbitrary values that are different from the unique value common to all the other views) on a subset of views. The set should not be large, otherwise the recognition process will fail too often.

Given a function f, for every object x let $E_f(x)$ denote the set of viewing directions for which f is incorrect ($E_f(x)$ is defined on the unit sphere). The object x again is taken to be a point in R^n. We also assume that objects that are very similar to each other have similar sets of "bad" viewing directions. For example, if the cube in Fig 2 is slightly distorted, the "bad" singular view will be only slightly shifted. More specifically, let us define for each object x, the value $\Phi(x, \epsilon)$ to be the measure (on the unit sphere) of all the viewing directions for which f is incorrect on at least one object in the neighborhood of radius ϵ around x. That is, $\Phi(x_0, \epsilon)$ is the measure of the set $\bigcup_{x \in B(x_0, \epsilon)} E_f(x)$. We can now show that even if $\Phi(x, \epsilon)$ is rather substantial (i.e. f makes errors on a significant number of views), f is still the trivial (constant) function. Specifically, assuming that for every x there exist an ϵ such that $\Phi(x, \epsilon) < D$ (where D is about 14% of the possible viewing directions), then f is a constant function. The proof of this claim is given in Appendix 2.

3.2 "Almost consistent" recognition functions

In practice, a recognition function may also not be entirely consistent in the sense that the function values for different images of the same object may not be identical, but only close to one another in some metric space (e.g., within an interval in R). In this case, a threshold function is usually used to determine whether the value indicates a given object.

Let an *object neighborhood* be the range to which a given object is mapped by such an "almost consistent" function. Clearly, if the neighborhood of an object does not intersect the neighborhoods of other objects, then the function can be extended to be a consistent function by a simple composition of the threshold function with the almost consistent function. In this case, the result of the general case (Claim 1) still holds, and the function must be the trivial function.

If the neighborhoods of two objects, a and b, intersect, then the scheme cannot discriminate between these two objects on the basis of images that are mapped to the intersection. In this case the images mapped to the intersection constitute a set of images for which f is inconsistent. If the assumption from the previous section holds, then f must be again the trivial function.

We have shown that an imperfect universal recognition function is still a constant function. It follows that any non model-based recognition scheme with a universal scope cannot discriminate between objects, even if it is allowed to make errors on a significant number of images.

4 Consistent recognition functions for class of objects

So far we have assumed that the scope of the recognition scheme was universal. That is, the recognition scheme could get as its input any set of (pointwise) 3-D objects. The recognition functions under consideration were therefore universally consistent with respect to viewing position. Clearly, this is a strong requirement. In the following sections we consider recognition schemes that are specific to classes of objects. The recognition function, in this case must still be consistent with respect to viewing position, but only for objects that belong to the class in question. That is, the function must be invariant to viewing position for images of objects that belong to a given class of objects, but can have arbitrary values for images of objects that do not belong to this class.

The possible existence of a nontrivial consistent recognition function for an object class depends on the particular class in question. In Section (4.1) we consider a simple

class for which a nontrivial consistent function (with respect to viewing position) still does not exist. In Section (4.2) we discuss the existence of consistent functions for certain infinite classes of objects. We show that when a nontrivial consistent function exist, the upper bound of any function discrimination power can be determined. Finally, we use the class of symmetric objects (Section 4.2.2) in order to demonstrate the existence of consistent function for an infinite class of objects and its discrimination power.

4.1 The class of a prototypical object

In this section, we consider the class of objects that are defined by a generic object. The class is defined to consist of all the objects that are sufficiently close to a given prototypical object. For example, it is reasonable to assume that all faces are within a certain distance from some prototypical face. For objects composed of n points in space, such a class can be thought of as a sphere in R^{3n} around the prototypical object.

The results established for the unrestricted case hold for such classes of objects. That is, every consistent recognition function with respect to viewing position of all the objects that belong to a class of a given prototypical object is a constant function. The proof for this case is similar to the proof of the general case in Claim 1.

4.2 A consistent recognition function

Nontrivial consistent recognition functions can be defined for many infinite classes of objects. For example, consider the infinite class of the eight-point objects with the points lying on the corners of some rectangular prism, together with the class of all three-point objects. Clearly, a nontrivial consistent function for this class can be defined. However, using the same proof as in Claim 1, it can be shown that such a function will have only two possible values, one for the three-point objects and the other for the eight-point objects. Hence, the function can be used for classification of three and eight-point objects, but cannot be used for identification of these objects. In this example the function is consistent for the class, all the views of a given object will be mapped to the same value. However, the function has a limited discrimination power, it can only distinguish between two subclasses of objects. In the next section we examine further the discrimination power of a recognition function.

4.2.1 Upper bound for function discrimination power

Given a class of objects, we first define a *reachability partition* of equivalence subclasses. Two objects are within the same equivalence subclass if and only if they are reachable

from each other. Reachability is clearly an equivalence relation and therefore it divides the class into equivalence subclasses. Every function f induces a partition into equivalent subclasses of its domain. That is, two objects, a and b, belong to the same equivalent subclass if and only if $f(a) = f(b)$. Every consistent recognition function must have the same value for all objects in the same equivalence subclass defined by the reachability partition (the proof is the same as in Claim 1). However, the function can have different values for images of objects from different subclasses. Therefore, reachability partition is a refinement of any partition induced by a consistent recognition function. That is, every consistent recognition function cannot discriminate between objects within the same reachability partition subclass.

The reachability subclasses in a given class of objects determines the upper bound on the discrimination power of any consistent recognition function for that class. If the number of reachability subclasses in a given class is finite, then it is the upper bound for the number of values in the range of any consistent recognition function for this class. In particular, it is the upper bound for the number of objects that can be discriminated by any consistent recognition function for this class. Note that the notion of reachability and, consequently, the number of equivalence classes, is independent of the particular recognition function. If the function discrimination power is low, the function is not very helpful for recognition but can be used for classification, the classification being into the equivalence subclasses.

In a non model-based recognition scheme, a consistent function must assign the same value to every two objects that are reachable within the scope of the scheme. In contrast, a recognition function in a model-based scheme is required to assign the same value to every two objects that are reachable within the set of objects that the function must in fact recognize. Two objects can be unreachable within a given set of objects but be reachable within the scope of objects. A recognition function can therefore discriminate between two such objects in a model-based scheme, but not in a non model-based scheme.

4.2.2 The class of symmetric objects

The class of symmetric objects is a natural class to examine. For example, schemes for identifying faces, cars, tables, etc, all deals with symmetric (or approximately symmetric) objects. Every recognition scheme for identifying objects belonging to one of these classes, should be consistent only for symmetric objects.

In the section below we examine the class of bilaterally symmetric objects. We will determine the reachability subclasses of this class, and derive explicitly a recognition function with the optimal discrimination power. We consider images such that for every point in the image, its symmetric point appears in the image.

Without loss of generality, let a symmetric object be $(0, p_1, p_2, ..., p_{2n})$, where $p_i = (x_i, y_i, z_i)$ and $p_{n+i} = (-x_i, y_i, z_i)$ for $1 \leq i \leq n$. That is, p_i and p_{n+i} are a pair of symmetric points about the $y-z$ plane for $1 \leq i \leq n$. Let $p_i^r = (x_i^r, y_i^r, z_i^r)$ be the new coordinates of a point p_i following a rotation by a rotation matrix R and scaling by a scaling factor s. The new x-coordinates are:

$$\begin{aligned} x_i^r &= s\,(x_i r_{11} + y_i r_{12} + z_i r_{13}) \\ x_{n+i}^r &= s\,(-x_i r_{11} + y_i r_{12} + z_i r_{13}) \end{aligned}$$

In particular, for every two symmetric points p_i and p_{n+i}, $x_i^r - x_{n+i}^r = 2s x_i r_{11}$. For every i the following ratios hold:

$$\frac{x_i^r - x_{n+i}^r}{x_1^r - x_{n+1}^r} = \frac{x_i}{x_1}$$

In the same manner it can be shown that for every i the following ratios hold::

$$\frac{y_i^r - y_{n+i}^r}{y_1^r - y_{n+1}^r} = \frac{y_i}{y_1}$$

It follows that the ratios between the distances of two pairs of symmetric points do not change when the object is rotated in space and scaled.

We will show that these ratios define a nontrivial partition of the class of symmetric objects to equivalence subclasses of unreachable objects. Let $d(p_i, p_j)$ be the distance between the points p_i and p_j. Define the function h by

$$h(0, p_1, p_2, ..., p_{2n}) = \left(\frac{d(p_2, p_{n+2})}{d(p_1, p_{n+1})}, \frac{d(p_3, p_{n+3})}{d(p_1, p_{n+1})}, .., \frac{d(p_n, p_{2n})}{d(p_1, p_{n+1})}\right)$$

Claim 2: Every two symmetric objects a and b are reachable if and only if $h(a) = h(b)$.

Proof:

Let $h(a) = h(b)$. We have to show that a and b are reachable by a symmetric sequence. That is, there exists a sequence of symmetric objects starting at a and ending at b such that every two successive objects have an orthographic projection in common. This is proved in Appendix 3.

Let $h(a) \neq h(b)$. We have to show that a and b are not reachable by a sequence of symmetric objects. Assume that there is a sequence of symmetric objects starting at

a and ending at b such that every two successive objects have a common orthographic projection. For every two successive objects, a_i and a_{i+1}, $h(a_i) = h(a_{i+1})$ because a_i and a_{i+1} have a common orthographic projection and h is independent of the viewing position. It follows that for every two objects, a_i and a_j, in the sequence connecting the objects a and b, $h(a_i) = h(a_j)$. This contradicts the assumption that $h(a_1) = h(a) \neq h(b) = h(a_n)$. □

It follows from this Claim that a consistent recognition function with respect to viewing position defined for all symmetric objects, can only discriminate between objects that differ in the relative distance of symmetric points.

5 Consistent recognition function for grey level images

So far, we have considered only binary images. In this section we consider grey level images of Lambertian objects that consist of n small surface patches in space (this can be thought of as sampling an object at n different points). Each point p has a surface normal N_p and a reflectance value ρ_p associate with it. The image of a given object now depends on the points' location, the points' normals and reflectance, and also on the illumination condition, that is, the level of illumination, and the position and distribution of the light sources.

An image now contains more information than before: in addition to the location of the n points, we now have the grey level of the points. The question we consider is whether under these conditions objects may become more discriminable then before by a consistent recognition function. We now have to consider consistent recognition functions with respect to both illumination condition and viewing position. We show that a nontrivial universally consistent recognition function with respect to illumination condition and viewing position still does not exists.

Claim 3: Any universally consistent function with respect to illumination condition and viewing position, that is defined on grey level images of objects consisting of n surface patches, is the trivial function.

In order to prove this claim, we will show that every two objects are reachable. That is, there exists a sequence of objects starting with the first and ending with the second object, and every successive pair in the sequence has a common image. A pair of objects has a common image if there is an illumination condition and viewing position such that

the two images (the points' location as well as their grey level) are identical. The proof is given in Appendix 4.

We conclude that the limitation on consistent recognition functions with respect to viewing position do not change when the grey level values are also given at the image points. In particular, it follows that a consistent recognition scheme that must recognize objects regardless of the illumination condition and viewing position must be model-based.

6 Model-based recognition schemes

In this section we show a model-based recognition scheme that is both consistent and nontrivial. Clearly, if every two objects in the set are reachable, then a nontrivial consistent recognition function does not exist. We therefore consider sets of objects in which at least some of the objects are not reachable by a sequence of objects from the set. For such a set of objects, it is possible to define a recognition function that (*i*) will be consistent, (*ii*) will have different values for objects that do not share a view. The definition of the function, in this case, depends strongly on the 3-D models of these objects. In order to construct the function, we use as an example the linear combination approach (Ullman & Basri 1989).

An object model in the linear combination scheme contains a number of images of a given object together with the correspondence between the image points. Every image of the object in question, taken from an arbitrary viewing position, can be expressed as the linear combination of the location of the corresponding points in the model images. Given an image and a candidate model, the first stage of the linear combination scheme consists of computing the coefficients of the linear combination. The linear combination of the model images is then computed, and the result (the transformed model) is compared with the image. If the two agree, the image is identified as an instance of the model. In the case of a finite set of objects, this computation can be repeated for every model in the set. The value of the function can be, for example, a canonical name for the object. Clearly, the function has the same value for all images of the same object and different values for images of different objects that do not share an orthographic projection.

Alternative schemes may also be used for the same task. The main point of the example is that a consistent recognition function that is as discriminating as possible, can be tailored to a given set of objects.

7 Conclusion

In this paper we have established some limitations on non model-based recognition schemes. In particular, we have established the following claims:

- Every function that is invariant to viewing position of all possible point objects is a constant function. It follows that every consistent recognition scheme must be model-based.
- If the recognition function is allowed to make mistakes and misidentify each object from a substantial fraction of viewing directions (about 14%) it is still a constant function.

We have considered recognition schemes restricted to classes of objects and showed the following: For some classes (such as classes defined by prototypical object) the only consistent recognition function is the trivial function. For other classes (such as the class of symmetric objects), a nontrivial recognition scheme exists. We have defined the notion of the discrimination power of a consistent recognition function for a class of objects. We have shown that it is possible to determine the upper bound of the function discrimination power for every consistent recognition function for a given class of object. The bound is determined by the number of equivalence subclasses (determined by the reachability relation). For the class of symmetric objects, these subclasses were derived explicitly.

For grey level images, we have established that the only consistent recognition function with respect to viewing position and illumination conditions is the trivial function.

In this study we considered only objects that consist of points on surface patches in space. Real objects are more complex. However, many recognition schemes proceed by first finding special contours or points in the image, and then applying the recognition process to them. The points found by the first stage are usually projections of stable object points. When this is the case, our results apply to these schemes directly. For consistent recognition functions that are defined on contours or surfaces, our result do not apply directly, unless the function is applied to contours or surfaces as sets of points. In the future we plan to extend the result to contours and surfaces.

Appendix 1

In this Appendix we prove that in the general case every two objects are reachable from one another.

First note that the projection of two points, when viewed from the direction of the vector that connects the two points, is a single point. It follows that for every object with $n-1$ points there is an object with n points such that the two objects have a common orthographic projection. Hence, it is sufficient to prove the following claim:

Claim 4: Any two objects that consists of the same number of points in space are reachable from one another.

Proof: Consider two arbitrary rigid objects, a and b, with n points. We have to show that b is reachable from a. That is, there exists a sequence of objects such that every two successive objects have a common orthographic projection.

Let the first object in the sequence be a. Each object in the sequence consists of the same points as the previous object, except for one point of object a which is replaced by a new point of object b. For example, the second object in the sequence consists of $n-1$ points of object a and one point of object b.

The formal definition of the sequence can be written as follows. Let the object a be $a = (p_1^a, p_2^a, ..., p_n^a)$ and the object b be $b = (p_1^b, p_2^b, ..., p_n^b)$ where p_i^a and p_i^b are points in 3-D. The first and last objects in the sequence are a and b, i.e., $a_1 = a$ and $a_{n+1} = b$. We take the rest of the sequence, $a_2, ..., a_n$ to be the objects: $a_i = (p_1^b, p_2^b, ..., p_{i-1}^b, p_i^a, ..., p_n^a)$. For example $a_2 = (p_1^b, p_2^a, ..., p_n^a)$.

The first and the last objects in the sequence are a and b, respectively. All that is left to show is that for every two successive objects in the sequence there exists a direction such that the two objects project to the same image. By the sequence construction, every two successive objects differ by only one point. The two non-identical points project to the same image point on the plane perpendicular to the vector that connects them. Clearly, all the identical points project to the same image independent of the projection direction. Therefore, the direction in which the two objects project to the same image is the vector defined by the two non-identical points of the successive objects. □

Appendix 2

In this appendix we show that even an imperfect recognition function is a constant function, provided that the sets of viewing directions on which it fails are not too large.

An object x is taken to be a point in R^n. For each object x we assume that f gets a unique value, considered the "correct" value for x, for most of the views, but it is allowed to have different, incorrect values, for other views. $\Phi_f(x, \epsilon_a)$ is the measure of the set of viewing directions for which f is incorrect on at least one object in a neighborhood of

radius ϵ_x around x (the units of $\Phi_f(x, \epsilon_x)$ are Steradians[2]). We wish to show that if for every x there exists an ϵ_x such that $\Phi_f(x, \epsilon_x) < D$ for a certain constant D, then f must be the trivial (constant) function. We give a lower bound on D which is about 14% of the set of all viewing directions.

Let us define a pair of objects, (a, b), to be an *f-correct* pair if a and b have a common orthographic projection and the value of f on this common orthographic projection is correct for both objects. An *f-correct* sequence is a sequence such that each successive pair of objects is an f-correct pair. We say that objects a and b are *f-reachable* if there exists an f-sequence starting at a and ending at b.

Using a similar proof to the general case (Claim 1), it is sufficient to prove the following Lemma.

Lemma: Let f be a recognition function defined on objects that consist of n points in space. Let $\Phi_f(x, \epsilon_x)$ be the measure of viewing directions for which f is incorrect on at least one of the objects in the ϵ_x neighborhood of x. Assume that for every x there exists an ϵ_x such that $\Phi_f(x, \epsilon_x) < D$ (D is fixed for all objects and taken below 0.92 Steradians which is about 14% percent of all possible viewing directions). Then, every two objects (consisting of n points in space) are f-reachable.

Proof: Every object that consists of n points in space can be regarded as a point in R^{3n}. We assume that, if f is correct on one image, then it will also be correct on the same image scaled by any factor. That is, if a and b are f-reachable, then a and b scaled by any factor are also f-reachable. Hence, it is sufficient to consider only objects that are points in the unit sphere in R^{3n}, which we denote by B_0^{3n}.

Let a and b be two objects in B_0^{3n}. Consider the original sequence of objects connecting a and b as in Claim 1 (see Appendix 1). If the sequence is f-correct, then we are done. Otherwise, using the three claims below we will show that it is always possible to construct an f-correct sequence of objects connecting any f-incorrect pair of objects (note that successive objects in the sequence differ by a single point). The f-correct sequence between the objects a and b is, then, the original sequence with additional subsequences between all the originally f-incorrect pairs of objects. We first list the three claims, then give their proofs.

Claim 5: There exists a fixed r such that for every object $x \in B_0^{3n}$, $\Phi_f(x, r) < D$.

In the following two claims, let (a_i, a_{i+1}) be an f-incorrect pair of successive objects from the original sequence. Let d be the distance between a_i and a_{i+1} as measured in R^{3n}.

[2]A Steradian is the area cut out by a cone of directions on the surface of the unit sphere surface. The area cut out by a cone with apex angle α on the surface of a unit sphere is: $2\pi(1 - cos(\alpha))$.

Claim 6: If $d < 2r$ then there exists an object c_0 such that the pairs (a_i, c_0) and (a_{i+1}, c_0) are f-correct.

Claim 7: If $d \geq 2r$, then there exist two objects, c_0 and c_1 such that:
(a) The pairs (c_0, a_i) and (c_1, a_{i+1}) are f-correct.
(b) The distance between c_0 and c_1 is less than $d-\rho$, where ρ is a constant strictly greater than zero.

We now show that these claims suffice. Given the original sequence from Claim 1, replace every f-incorrect pair of objects (a_i, a_{i+1}) such that $d < 2r$ by the subsequence (a_i, c_0, a_{i+1}) from Claim 6. Replace every f-incorrect pair of objects (a_i, a_{i+1}) such that $d \geq 2r$ by the subsequence (a_i, c_0, c_1, a_{i+1}) from Claim 7. If (c_0, c_1) is still f-incorrect, repeat the process until the distance between the two new objects is less than $2r$. Claim 7b guarantees that this process needs to be repeated only a finite number of times. As a result, an f-correct sequence consisting of finite number of objects is obtained.

We next prove claims $5 - 7$ above.

Proof of 5: Let $\bar{B}_0^{3n}$ be the close unit sphere in R^{3n}. For every $x \in \bar{B}_0^{3n}$ there exists an ϵ_x such that $\Phi_f(x, \epsilon_x) < D$. Consider the family of open sets $B^{3n}(x, \frac{\epsilon_x}{2})$ for every $x \in \bar{B}_0^{3n}$. This is an infinite cover of the unit sphere $\bar{B}_0^{3n}$. Since the sphere $\bar{B}_0^{3n}$ is a compact set, there exists a finite subset, $\{B^{3n}(x_i, \epsilon_i)\}_{i=0}^{m}$ that covers $\bar{B}_0^{3n}$. Let r be the minimum radius in the finite cover.

Every point $x \in B_0^{3n}$ satisfies $x \in B^{3n}(x_i, \frac{\epsilon_i}{2})$ for some $0 \leq i \leq m$. We thus have:

$$\Phi_f(x, r) \leq \Phi_f(x_i, 2r) \leq \Phi_f(x_i, 2\epsilon_i) \leq \Phi_f(x_i, \epsilon_{x_i}) < D$$

Hence, $\Phi_f(x, r) < D$ for every $x \in B_0^{3n}$. □

Proof of 6: By the sequence construction, the objects a_i and a_{i+1} differ by only one point. Let a be the object that consists of the $n-1$ identical points of a_i and a_{i+1}. Let p_i and p_{i+1} be the non-identical points of a_i and a_{i+1} respectively. We define the object $a \oplus b$ to be the object that consists of the points of both a and b. For example, $a_i = a \oplus p_i$ and $a_{i+1} = a \oplus p_{i+1}$. Let us define the function f' on images of one point objects in the following way: $f'(p) = f(a \oplus p)$.

Every object that consists of one point in space can be regarded as a point in R^3. Therefore, in order to consider the objects of the form $a \oplus p$, it is sufficient to consider the unit sphere in R^3 and the function f'.

Let p be the point $\frac{p_i + p_{i+1}}{2}$. The distance between p_i and p_{i+1} is less than $2r$. Therefore, $p_i, p_{i+1} \in B(p, r)$ (a ball of radius r centered at p"). Consider the plane v of equidistant

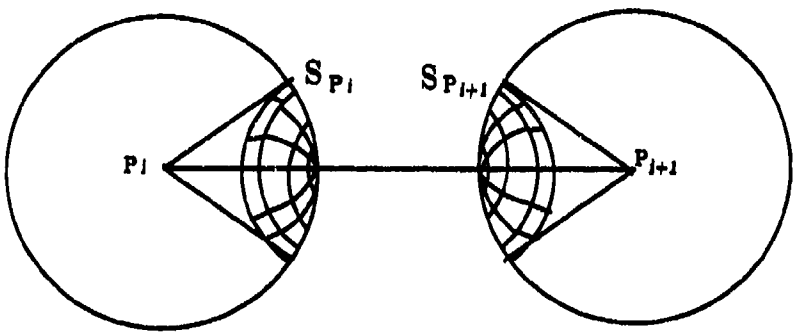

Figure 3: The two non-identical points, p_i and p_{i+1} and the portions of the spheres surfaces, S_{p_i} and $S_{p_{i+1}}$.

points from p_i and p_{i+1} in the sphere $B(p,r)$. We claim that there exists a point p_c on v such that the pairs (p_i,p_c) and (p_{i+1},p_c) are f' correct. Let us assume that such a point does not exist. That is, for every point $p_c \in v$ one of the pairs (p_c,p_i) or (p_c,p_{i+1}) is f-incorrect. The function f must, therefore, be incorrect on at least D directions of viewing position for one of the objects in $v \cup \{p_i,p_{i+1}\}$. It follows that $\Phi_{f'}(p,r) \geq D$. From the definition of Φ and f' it follows that for every $p \in B_0^3$, $\Phi_{f'}(p,r) \leq \Phi_f(a \oplus p,r)$. We can assume that without loss of generality[3] the object $a_p = a \oplus p$ is in B_0^{3n}. If this is the case it can be shown that a_p is in B_0^{3n}. By Claim 1 we have that $\Phi_f(a \oplus p,r) < D$ and hence also $\Phi(p,r) < D$. Hence we have a contradiction. □

Proof of 7: Let p_i and p_{i+1} be the non-identical points of a_i and a_{i+1}. Consider the two spheres $B_i = B(p_i,r)$ and $B_{i+1} = B(p_{i+1},r)$. Let S_{p_i} be the portion of the sphere surface $B(p_i,r)$ cut by a cone whose apex is p_i, whose axis is the vector $p_i - p_{i+1}$, and apex angle 35° (see Fig 3). $S_{p_{i+1}}$ is defined in a similar manner.

There must be at least one point, p_{c_0} on S_{p_i} such that the pair (p_i,p_{c_0}) is f'-correct. Otherwise, $\Phi_{f'}(p_i,r) \geq D$ in contradiction with the fact that $\Phi_f(a \oplus p_i,r) = \Phi_f(a_i,r) < D$ (Claim 5). For the same reason, there exists a point p_{c_1} on $S_{p_{i+1}}$ such that the pair (p_{i+1},p_{c_1}) is f' correct. All that is left to be shown is that the distance, s, between p_{c_0} and p_{c_1} is smaller than $d - \rho$.

Let s_m be the maximal distance between two points on S_{p_i} and $S_{p_{i+1}}$ for a given d. It is sufficient to show that $d - s_m \geq \rho$ (note that s is a function of d). Let $p_{c_0} \in S_{p_i}$ and $p_{c_1} \in S_{p_{i+1}}$ be two points such that the distance between them is maximal for a given d. From symmetry considerations the line connecting the points p_{c_0} and p_{c_1} intersects the

[3] by scaling the two given objects such that the distance between each of them and the unit sphere surface will be at least r

line that connects the points p_i and p_{i+1}. Denote by o the intersection point. Denote by s_0 and s_1 the distances between p_{c_0} and o, and p_{c_1} and o respectively, denote by d_0 and d_1 the distances between p_i and o and p_{i+1} and o respectively. Using the cosine rule, $s_0 = \sqrt{d_0^2 - 2d_0 r\cos(\alpha) + r^2}$, where $0 \leq \alpha \leq 35$. Clearly, s_0 has maximum when $\alpha = 35$.

s_0 is monotonically increasing in d (it can be readily proved by taking the derivative $\frac{ds_0}{dd}$). Therefore it is sufficient to compute s_r, the value of s_0 when $d = r$. In that case, $s_r = r\sqrt{2 - 2\cos(35)}$. From symmetry considerations, $s_0 = s_1$ and $d_0 = d_1$. Therefore we obtain, $d - s \geq d - s_m = 2(d_0 - s_0) \geq 2(d - s_r) \geq 2(r - s_r) \geq 2r(1 - \sqrt{2 - 2\cos(35)}) = \rho$.

Appendix 3

In this appendix we prove that every two symmetric objects, a and b such that $h(a) = h(b)$, are reachable by a sequence of symmetric objects.

Let $a = (0, p_1^a, p_2^a, ..., p_{2n}^a)$ and $b = (0, p_1^b, p_2^b, ..., p_{2n}^b)$. Let the first object in the sequence be a. The second object in the sequence is the object a scaled by $\frac{d(p_1^b, p_{n+1}^b)}{d(p_1^a, p_{n+1}^a)}$. Denote the second object in the sequence by a'. By our assumption $h(a) = h(b)$, that is, $\frac{d(p_1^a, p_{n+1}^a)}{d(p_1^a, p_{n+1}^a)} = \frac{d(p_1^b, p_{n+1}^b)}{d(p_1^b, p_{n+1}^b)}$. The projection of a' and b on the $x \times y$ plane are symmetric images that satisfy, $x_i^{a'} = x_i^b$ for every i.

Each object in the sequence consist of the same points as its previous object except for a pair of symmetric points of the object a' which is replaced by a new pair of symmetric points of the object b. The direction for which the two objects project to the same image is the vector that connects one point from the pairs of symmetric point from a' with one point of symmetric point from b. Note that this vector is in the $y - z$ plane, hence the symmetry of the image is kept.

Appendix 4

In this appendix we show that every two objects a and b composed of n small surface patches are reachable by a sequence of objects starting with a and ending with b such that for each successive pair of objects there exists a common image. It will then follow that a nontrivial consistent recognition function with respect to illumination condition and viewing position does not exist.

The grey level at each image point is determined by the illumination condition, the normal direction and reflectance value of the object point. Given two n-point objects,

a and b, we construct the same sequence as in Claim 1, but the normal direction and reflectance value of each new b point (that replaces an a point) is taken to have the normal direction and reflectance value of the corresponding point from object a. In this manner we get a sequence of objects starting at a and ending at some b' such that each pair of successive objects have a common image (the same location and grey level values of the image points). The objects b and b' have identical configurations of n points in space, but with possibly different normal directions and albedo values associated with corresponding points. Therefore, the object b and b' do not necessarily have a common image.

It is left to show, then, that every two n-point objects having the same configuration in space with possibly different normal direction and albedo values, are reachable. That is, there exists a sequence of objects, and for every two successive objects there is an illumination condition, such that the projected grey level of the two objects is identical at every point.

We construct a sequence such that the first and the last objects in the sequence are b' and b respectively, and every two successive objects differ at only one point. Let $p_{b'}$ and p_b be the two non-identical points in a successive pair. Let $\hat{N}_{b'}$ and $\hat{N}_b$ be the unit vectors in the normal directions, $\rho_{b'}$ and ρ_b be the albedo of the points $p_{b'}$ and p_b respectively. We will assume first that the objects are Lambertian (for details regarding the images of Lambertian surfaces see Horn 1977). In this case, the intensity at the points is given by $I_{b'} = \rho_{b'} E \cdot \hat{N}_{b'}$ and $I_b = \rho_b E \cdot \hat{N}_b$, where E is the light source vector (its direction in pointing at the equivalent light source, and its magnitude is proportional to the source intensity). For the two objects to have a common image, it is sufficient to find an illumination vector E such that the $p_{b'}$ and p_b will have identical grey levels, i.e. $\rho_b \cdot E \cdot \hat{N}_b = \rho_{b'} \cdot E \cdot \hat{N}_{b'}$. Such an E clearly exists, because it is defined by one linear equation in three variables. The vector E should also satisfy $E \cdot \hat{N}_{b'} > 0$ and $E \cdot \hat{N}_b > 0$. This is clearly possible since if $E \cdot \hat{N}_b < 0$, then $E \cdot \hat{N}_{b'} < 0$ as well, and we can choose $-E$ for our final solution[4]

In case the object is not Lambertian but has a specular component, the intensity at each point becomes the sum of the Lambertian component and the specular component. The intensity value due to the specular component depends on the viewing position, the surface normal, the light source position, and some other surface specular parameters (Phong 1975). We have not considered this case in detail but it seems that by choosing the light source and the viewing position, two successive objects in the sequence still have a common image.

[4]In case that $\hat{N}_{b'} = \hat{N}_b$ but $\rho_{b'} \neq \rho_b$, the solution for E is such that $E \cdot \hat{N}_{b'} = 0$. We can than add one intermediate object to the sequence, with albedo ρ_b and normal $\hat{N}'_{b'} \neq \hat{N}_{b'}$.

References

- Bolles, R.C. and Cain, R.A. 1982. Recognizing and locating partially visible objects: The local-features-focus method. *Int. J. Robotics Research*, 1(3), 57-82 .

- Brooks, R.A. 1981. Symbolic reasoning around 3-D models and 2-D images, *Artificial Intelligence J.*, 17, 285-348.

- Burns, J. B., Weiss, R. and Riseman, E.M. 1990. View variation of point set and line segment features. *Proc. Image Understanding Workshop, Sep.*, 650-659.

- Cannon, S.R., Jones, G.W., Campbell, R. and Morgan, N.W. 1986. A computer vision system for identification of individuals. *Proc. IECON 86 0, WI.*, 1, 347-351.

- Clemens D.J. and Jacobs D.W. 1990. Model-group indexing for recognition. *Proc. Image Understanding Workshop, Sep.*, 604-613.

- Cyganski, D., Cott, T.A., Orr, J.A. and Dodson, R.J. 1987. Development, implementation, testing and application of an Affine transform invariant curvature function. *Proceeding of ICCV Conf., London*, 496-500.

- Grimson, W.E.L and Lozano-Pérez, T. 1984. Model-based recognition and localization from sparse data. *Int. J. Robotics Research*, 3(3), 3-35.

- Grimson, W.E.L and Lozano-Pérez, T. 1987. Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. on PAMI.* 9(4), 469-482.

- Horn B. K.P. 1977. Understanding image intensities, *Artificial Intelligence J.*. 8(2), 201-231

- Hu, M. K., 1961. Pattern recognition by moments invariants. Proc *IRE* 49, 1428.

- Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Trans. Inform. Theory. **IT-8**, 179-187.

- Huttenlocher, D.P. and Ullman, S. 1987. Object recognition using alignment. *Proceeding of ICCV Conf., London*, 102-111.

- Kanade, T. 1977. Computer recognition of human faces. *Birkhauser Verlag. Basel and Stuttgart.*

- Khotanzad, A. and Hong, Y.H. 1990. Invariant image recognition by Zernike moments. *IEEE trans. on PAMI, 12*(5), 489-497.

- Lamdan, Y. and Wolfson, H. J. 1988. Geometric hashing: A general and efficient model-based recognition scheme. *Proceeding of ICCV Conf., Tampa, Florida*, 238-249.

- Lin, C. 1987. New forms of shape invariants from elliptic Fourier descriptions. *Pattern Recognition*, **20**(5), 535-545

- Lowe, D.G. 1985. Three dimensional object recognition from single two-dimensional images. *Robotics research Technical Report 202, Courant Inst. of Math. Sciences, N.Y. University.*

- Phong, B.T. 1975. Illumination for computer generated pictures. *Communication of the ACM*, **18**(6), 311-317.

- Poggio T., and Edelman S. 1990. A network that learns to recognize three dimensional objects. *Nature*, **343**, 263-266.

- Ponce, J., Chelberg, D. and Mann, W. 1985. Invariant properties of the projection of straight homogeneous generalized cylinders. *Proceeding of ICCV Conf., London*, 631-635.

- Ullman S. 1977. Transformability and object identity. *Perception and Psychophysics*, **22**(4), 414-415.

- Ullman S. 1989. Alignment pictorial description: an approach to object recognition. *Cognition*, **32**(3), 193-254.

- Ullman S. and Basri R. 1989. Recognition by linear combinations of models *AI MEMO No. 1152, AI MEMO No 1152, The Artificial Intelligence Lab., M.I.T.*

- Verri A. and Yuille A, 1986. Perspective projection invariants. *AI MEMO No. 832, The Artificial Intelligence Lab., M.I.T.*

- Wong, K.H., Law, H.H.M. and Tsang P.W.M, 1989. A system for recognizing human faces, *Proc. ICASSP*, 1638-1642.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1991	memorandum

4. TITLE AND SUBTITLE
Limitations of Non Model-Based Recognition Schemes

5. FUNDING NUMBERS
N00014-86-K-0685
DACA76-85-C-0010
N00014-85-K-0124
IRI-8900267

6. AUTHOR(S)
Yael Moses and Shimon Ullman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER
AIM 1301

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Information Systems
Arlington, Virginia 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution of this document is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Abstract: Different approaches to visual object recognition can be divided into two general classes: model-based vs. non model-based schemes. In this paper we establish some limitation on the class of non model-based recognition schemes. A non model-based scheme is based on functions invariant to viewing position and illumination conditions. We show that every function that is invariant to viewing position of all objects is the trivial (constant) function. The same result holds even if the recognition function is not required to be perfect, but is allowed to make mistakes and misidentify each object from a substantial fraction of viewing directions. It follows that every consistent recognition scheme for recognising 3-D objects must in general be model based.

(continued on back)

14. SUBJECT TERMS (key words)
recognition
object recognition
invariant properties
model-based vision

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18
298-102

Block 13 continued:

We then consider recognition schemes restricted to classes of objects and show that, for some classes, the only consistent recognition function is still the trivial function. For other classes (such as the class of symmetric objects) a nontrivial recognition scheme exists. We define the notion of a discrimination power of a consistent recognition function for a class of objects. The function's discrimination power determines the set of objects that can be discriminated by the recognition function. We show that it is possible to determine the upper bound of the function's discrimination power for every consistent recognition function.